Calculating Areal Ranfall Using A More Efficient Idw Interpolation Algorithm

Peter Valent[1], Roman Výleta[2]

[1,2]Department of Land and Water Resources Management, Faculty of Civil Engineering, Slovak University of Technology, Slovakia

***Abstract*— *The estimation of areal rainfall is an important part of solving various hydrological problems utilizing rainfall-runoff and other models. Traditionally used Thiessen polygons (TP) method proved to be inaccurate mainly in mountainous areas and in catchments with insufficient number of rainfall gauges. One of the alternative to this method is an inverse distance weighting (IDW) method giving better estimates of areal rainfall even on places with rugged orography. However, this and similar methods are far less efficient than the simple TP method restraining its use to tasks where computational efficiency is not important. In this study a new algorithm accelerating the traditional IDW method is proposed and applied to three mountainous catchments situated in the central part of Slovakia. The method is compared with traditional IDW and TP methods in terms of both computational efficiency and estimated values. The results showed that while the new method gives the same results as the traditional IDW method it is far more efficient when the computational time was in all three catchments reduced by more than 96%.*

Keywords - areal rainfall, inverse distance weighting, interpolation, Thiessen polygons.

I. INTRODUCTION

In hydrology and other water related disciplines rainfall is one of the most important variables determining the amount of water entering the system. Rainfall measurements are essential for solving a wide range of tasks including those in meteorology, hydrology, agriculture, climate research or hydropower [7]. Even though the recent years and decades gave rise to very sophisticated measurement techniques such as meteorological radars or satellites most of the data is still collected with elevated can-type rainfall gauges. These enable to collect precipitation amounts only at single sites. However, rainfall data have very strong spatial variability and thus it is not reasonable to presume that a larger area around the rainfall gauge will share the same precipitation amounts. In order to account for the spatial variability a network comprising of several rain gauges needs to be deployed to correctly represent the rainfall amounts over a certain area. Generally, the denser the network is the more accurate is the information about the distribution of rainfall amounts over a certain area. Over the decades a large number of models of various complexity have been developed to utilize this information to solve various tasks in hydrology and other disciplines. Some of these tasks are of high importance such as predicting floods, managing water resources, evaluating crop yields and development or assessing the impact of land use or climate changes on the runoff from a catchment. Most of these models are lumped dealing with the catchment as a single unit and thus require estimates of mean areal rainfall as model input. The estimation of areal rainfall from a number of rainfall gauges could be performed using a number of various techniques. The overview of these techniques and their comparison is given in [1], [5], [8] or [9].

One of the most conventional and traditional one is the Thiessen polygon technique [10]. This method is favoured due to its simplicity, ease of use and minimum requirements on computational power. However, several authors (see e.g. [6], [8], [2] and [3]) comparing the Thiessen polygon technique with other techniques concluded that due to its fundamental principles it may produce inaccurate results especially on places with high topographical variation and the limited number of available rainfall gauges [9]. One of the possible issues with the Thiessen polygon method is displayed in (Fig. 1a), where the rainfall gauge situated in the upper right corner has no influence on the mean areal rainfall even though it is in the very near vicinity of the catchment.

Another approach in calculating areal rainfall uses inverse distance weighting (IDW) technique belonging to a family of distance weighting techniques ([8] and [4]). This technique eliminates the drawback of the Thiessen polygon method by dividing the catchment into a grid and interpolating rainfall amounts into each cell of the grid. The technique utilizes data from all rainfall gauges while assigning higher weights to stations closer to the interpolated cells. Fig. 2b shows that using the IDW technique the upper right station effects the mean areal rainfall over the catchment of interest. One of the disadvantages of the method used in calculating areal rainfall is that the precipitation must be interpolated into each cell of the grid lying inside the catchment boundaries. Even though the interpolation into particular cells is not computationally

demanding in the case of large catchments or dense grids the calculation is repeated numerous times (in case of 500×500 grid it is 250 000 times) for each day (or other time step) for which the areal rainfall is calculated. This produces a large number of interpolated maps which makes it computationally very demanding (in case of calculating daily areal rainfall amounts for 10 years this would lead to more than 900 million interpolations).

When calculating areal rainfall the information about the spatial variability of the precipitation is redundant and not needed. By avoiding the interpolation of rainfall amounts into each cell of the grid the computational requirements would decrease significantly. In this study an algorithm for accelerating calculation of mean areal rainfall using the IDW method was proposed. The new algorithm was compared with traditional IDW method and with the Thiessen polygon method in terms of calculation duration and estimated values.

Fig (A) Fig (B)

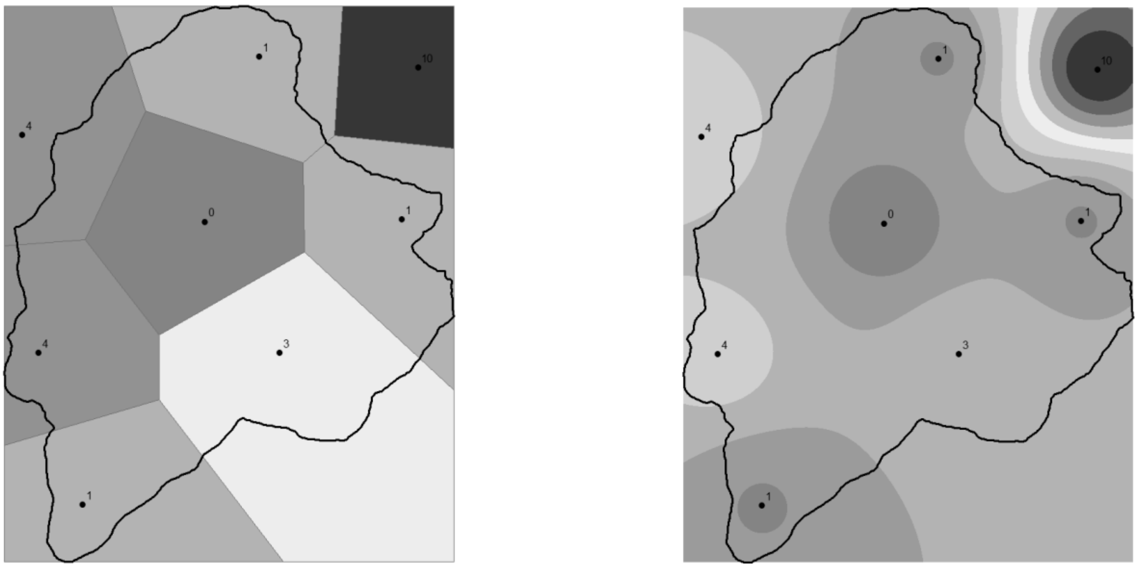

FIG. 1: AN EXAMPLE OF CALCULATING A REAL RAINFALL USING TWO INTERPOLATION TECHNIQUES: A) THIESSEN POLYGONS AND B) IDW.

II. METHODOLOGY

Areal rainfall over a certain area *A* is commonly defined using the following equation:

$$P_t^A = \frac{1}{A}\int_A P_t(x,y)\,dx\,dy \tag{1}$$

where $P_t(x,y)$ denotes the rainfall depth at the point (x,y), for the tth time interval. The total rainfall P_t^A over the area *A* is unknown, since the rainfall depth is accessible only at finite number of locations *n* (rainfall gauges). In hydrology the areal rainfall is then estimated using so called linear estimators given by

$$\hat{P}_t^A = \sum_{i=1}^{n} w_i P_i^t \tag{2}$$

which is a weighted mean of the random variables $P_t^1, P_t^2, \ldots, P_t^n$ observed at the rainfall gauges with weights w_i, which sum is always 1. Both Thiessen polygons and IDW techniques are linear estimators differing from one another in the values of weights w_i [5].

A. Thiessen polygons method

In this method the catchment is divided into *n* zones of influence A_i delineating an area which inner points' closest station is station *i*. The weighting coefficients w_i represent the proportion of particular zones of influence A_i to the total area of the

catchment *A*. The weighting coefficients w_i are then computed as

$$w_i = \frac{A_i}{A} \quad i = 1, 2, \ldots, n \tag{3}$$

The areal rainfall is then estimated from equation (2) which is a sum of weighted rainfall depths from all observed gauges. In the case of missing data in one or more gauges the weights had to be recalculated and areal rainfall was estimated from the remaining stations.

B. Inverse distance weighting method

The IDW interpolation method differs from the Thiessen polygons method in the way the weights are calculated. In the IDW method the weights are solely a function of the distances between the point of interest (x_0, y_0) and the rainfall gauges (x_i, y_i). This relationship is given by:

$$w_i = \frac{f(d_{0i})}{\sum_{i=1}^{n} f(d_{0i})} \tag{4}$$

where d_{0i} is the distance between the point of interest and the ith rainfall gauge. Function $f(d_{0i})$ is a distance function given by:

$$f(d_{0i}) = \frac{1}{d_{0i}^{\beta}} \tag{5}$$

where β is an appropriate power constant usually from an interval $\mathbb{R}\langle 1, 3\rangle$. Equation (5) indicates that the higher the power function is the lower weight is assigned to remote gauges. In this study the power constant was set to be 2.

When interpolating rainfall depths over large areas or with network of rainfall gauges the information contained in remote stations could be neglected since it does not add any relevant additional information to the interpolation procedure (remote stations would have negligible weights and thus minor influence on the interpolated value and at places with a sufficient amount of observations no additional information is needed). Because of this most of the GIS software contain an additional setting limiting the number of stations included to the interpolation either by setting the maximum number of stations or the maximum distance from the point of interest (see Fig. 2).

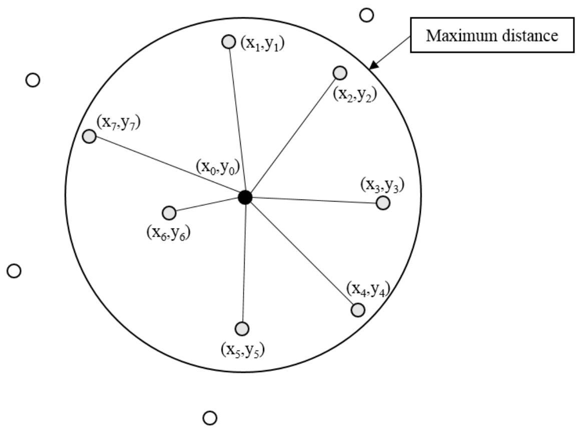

FIG. 2: AN EXAMPLE OF SELECTING STATIONS COMPLYING WITH THE MAXIMUM DISTANCE FROM THE POINT OF INTEREST RULE IN THE IDW INTERPOLATION METHOD.

The estimation of areal rainfall using the traditional IDW technique comprises of several steps. In the first step the rainfall

depth is interpolated into all cells of the grid into which the catchment was divided. Since the grid is usually rectangular (rainfall depths are interpolated even for cells lying outside the catchment (see Fig. 1b)) only cells lying inside the catchment are selected. In the last step the mean areal rainfall is calculated by averaging rainfall depths from all cells lying inside the catchment (see Fig. 3). In the case of missing data the interpolation is performed from remaining rainfall gauges.

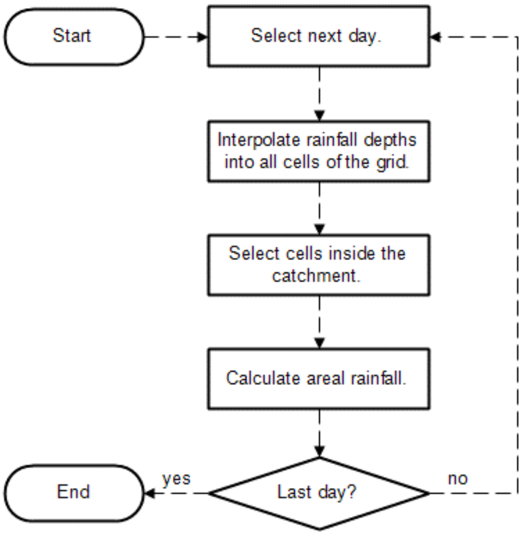

FIG. 3: SCHEME OF THE TRADITIONAL IDW INTERPOLATION TECHNIQUE

C. Accelerated IDW method

In this section a new algorithm for accelerating the calculation of mean areal rainfall using the IDW method is proposed. This method gives a simple 'shortcut' to the calculation of areal rainfall using the IDW interpolation technique which would otherwise have to be calculated using computationally demanding traditional approach (see the previous section). In order to use this technique let's assume that all rainfall gauges are used in interpolation (no maximum distance and number of stations limits).

Let $P'(x,y)$ be an interpolated rainfall depth at any arbitrary location with coordinates (x,y). This value can be calculated as a linear combination of observed values at stations P_1, P_2, $\cdots$, P_n and corresponding weights w_i $(i=1,2,\ldots,n)$. Rainfall depth at any location can be calculated by

$$P'(x,y)=\sum_{i=1}^{n} w_i P_i \tag{6}$$

Consider, the problem of finding the average within a catchment *A*. This average will be found by predicting $P'(x,y)$ separately for each grid cell within the catchment and averaging those predictions. Let the cell centres have coordinates $((u_1,v_1),\ldots,(u_m,v_m))$ where *m* is the number of cells in the region. The average $P'(A)$ is defined as

$$P'(A)=\frac{\sum_{j=1}^{m} P'(u_j,v_j)}{m} \tag{7}$$

By combining equations (6) and (7) and their subsequent modification we can get:

$$P'(A)=\sum_{i=1}^{n} P_i' \cdot \frac{\sum_{j=1}^{m} w(i,j)}{m}=\sum_{i=1}^{n} P_i' \cdot w_i' \tag{8}$$

where w_i' is only an average weight of a particular station over the whole catchment *A*. Let us recall that the individual

rainfall gauges contain a time series of observed rainfall depths. Let us denote rainfall depth at station i and time t P_{it}. Since the location of the gauges and the size of the grid cells is constant even the weights w_i' doesn't change in time. This implies that the areal rainfall in time t can be calculated as a linear combination of weights w_i' and observed rainfall depths P_{it}. Since the rainfall depths at stations P_{it} are known only the n weights w_i' have to be estimated. This can be done by setting value at station P_{it} at 1 and in all other stations at 0. After substituting these values to equation (8) we get

$$\begin{aligned} P'(A) &= w_1' \cdot 0 + \ldots + w_i' \cdot 1 + w_{i+1}' \cdot 0 + \ldots + w_n' \cdot 0 \\ P'(A) &= w_i' \end{aligned} \tag{9}$$

Equation (9) implies that the weights w_i' can be calculated as average values of synthetic rainfall depths (1 for i, 0 for others) over the catchment A. The calculation of the mean areal rainfall could be then reduced into the following steps:

Step 1: Set $P_{it} = 1$ and $P_{kt} = 0$ for $k \neq i$.

- Interpolate these values into each cell of the grid.
- Calculate mean of the interpolated values for each cell lying inside the catchment to get w_i'.

Step 2: Repeat step 1 for each station to get all weights w_i'.

Step 3: Calculate areal rainfall as a linear combination of observed rainfall depths P_{it} and weights w_i'.

Step 4: Repeat step 3 for each day with observations.

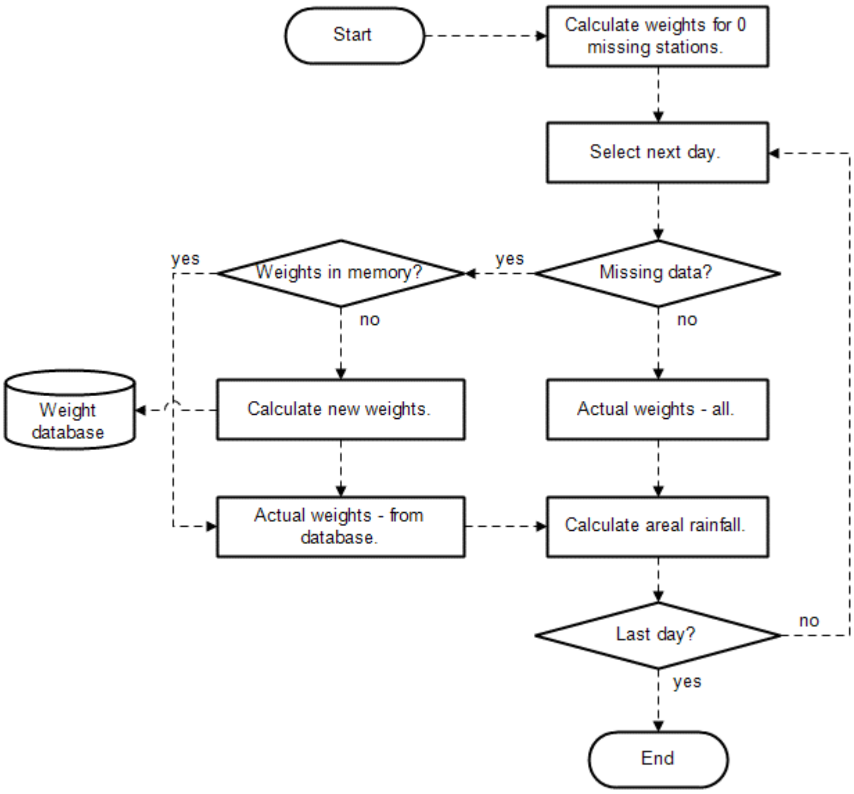

FIG. 4: SCHEME OF THE ACCELERATED IDW INTERPOLATION TECHNIQUE.

In the case of missing data in one or more stations it is necessary to recalculate the weights with the remaining stations (stations with missing data are not included into the interpolation). Since the same combination of missing stations could occur anywhere in the observed data (the same weights would have to be calculated several times) each combination of weights is stored in the weight database. If some of the stations at day *t* are missing the algorithm searches the database for the right combination of missing stations. If such combination is not found the new weights are calculated and stored in the database. The scheme of the algorithm is displayed in (Fig. 4).

III. DATA

The estimation of mean daily areal rainfall was tested on three mountainous catchments situated in the central part of Slovakia. The catchments are: 1) Hron at Banská Bystrica, 2) Hron at Brezno and 3) Váh at Liptovký Mikuláš (for the position of the catchments see Fig. 5). In all the catchments daily rainfall data were collected for the period between 1.1.1961 and 31.12.2010.

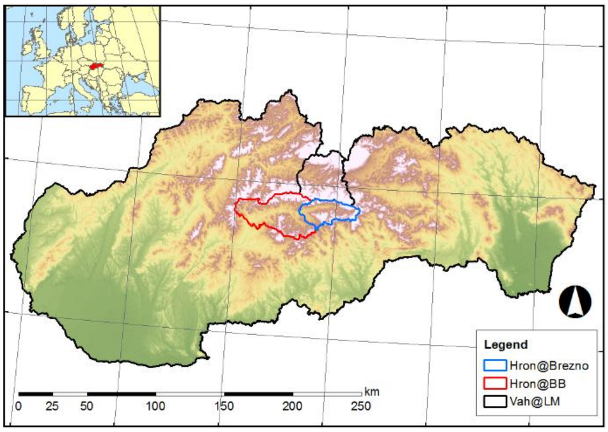

FIG. 5: SITUATION OF CATCHMENTS OF INTEREST IN THE CENTRE OF SLOVAKIA.

Catchment Hron with the outlet at Banská Bystrica (Hron@BB) has an area of 1768.2 km^2. The catchment has a shape of a narrow valley with two high mountains situated in the northern and southern part of the catchment. The average altitude of the catchment is 845 m ASL with minimum and maximum point at 350 and 2043 mASL respectively. The areal rainfall was estimated from 24 rainfall gauges situated inside the catchment and in its near vicinity.

Catchment Hron with the outlet at Brezno (Hron@Brezno) is a nested catchment of Hron@BB. It has an area of 578.6 km^2. Its mean altitude is 916 m ASL with minimum and maximum elevation of 489 and 1946 m ASL respectively. Rainfall depths were collected from 11 rainfall gauges situated mostly in the lower parts of the catchment.

Catchment Váh with the outlet at Liptovský Mikuláš (Vah@LM) is a neighbouring catchment to the Hron@BB with which it shares the southern border. Its area is 1100.6 km^2 with the mean altitude of 1090 m ASL (min. 568 m ASL and max. 2494 m ASL). Rainfall depths were collected from 28 rainfall gauges with some of the gauges situated in the highest parts of the catchment.

In all three catchments the rainfall datasets contain a certain percentage of missing data. The percentage of days with at least one missing value in any station and the percentage of missing values from all data is described in (Table 1).

TABLE 1
BASIC CHARACTERISTICS OF SELECTED CATCHMENTS. THE LAST TWO COLUMNS DESCRIBE THE PERCENTAGE OF DAYS WITH AT LEAST ONE MISSING DATA IN ANY STATION AND PERCENTAGE OF ALL MISSING VALUES.

Catchment	Area [km^2]	Mean alt. [m ASL]	Number of gauges	Period	Days with missing data	Missing values
Hron@BB	1768.2	845	24	1961 - 2010	83%	9.1%
Hron@Brezno	578.6	916	11		62%	9.6%
Vah@LM	1100.6	1090	28		77%	29%

IV. RESULTS AND DISCUSSION

In order to assess the performance of the new algorithm (IDW-a) the calculation of mean areal rainfall was compared with the traditional IDW method (IDW-t) and with the Thiessen polygons method (TP). The methods were compared both in terms of the estimated areal rainfall and their computational requirements.

A. Comparison of estimated areal rainfall

Since the main objective of the study was to propose a method which would accelerate the calculation of areal rainfall using the tradional IDW method it was crutial to demonstrate that the new method gives identical results. The method of Thiessen polygons was also included into the comparison to prove that it gives different estimates of areal rainfall. The individual methods were compared to the traditional IDW method (IDW-t) which was used as a reference method. The comparison showed that the difference between the estimates of areal rainfall calculated using the IDW-a and the IDW-t methods are negligible and thus could be considered as identical (see Table 2). When comparing the TP to the reference IDW-t method it is possible to observe a small difference between these two methods. Even though the maximum absolute difference between estimeted areal rainfall is almost 4 mm and in the case of Vah@LM more than 7 mm the mean absolute error is in all catchments lower that 0.2 mm. This means that even though the TP method doesn't usually give very accurate estimates in this particular case the estimates are comparable to those calculated by the IDW methods.

TABLE 2
THE COMPARISON OF SELECTED STATISTICS BETWEEN TP AND IDW-T METHODS AND IDW-A AND IDW-T METHODS

Catchment	Hron@BB		Hron@Brezno		Vah@LM	
	TP	IDW-a	TP	IDW-a	TP	IDW-a
Max abs. err.	3.9	<0.001	3.8	<0.001	7.3	<0.001
MAE	0.111	<0.001	0.123	<0.001	0.187	<0.001
RMAE	0.034	<0.001	0.34	<0.001	0.059	<0.001

B. Comparison of algorithm efficiency

The computational efficiency of the methods in estimating mean areal rainfall was tested at two datasets for each catchment. The first dataset comprised of 50 years of daily rainfall depths observed at individual rainfall gauges during the period between 1961 and 2010. This dataset was not complete and contained a certain percentage of missing data in particular stations (see Table 1). The computational efficiency of the methods was evaluated based on the time required for the estimation of areal rainfall in all days of the dataset. The results of the analysis are shown in Fig. 6 and Table 3. In all three catchments the slowest method is IDW-t method followed by IDW-a and TP method. In the Hron@BB catchment the time required for the estimation of areal rainfall using the traditional IDW-t method is over 5 hours and 43 minutes. The identical values (see Table 2) of areal rainfall were estimated using the proposed accelerated IDW-a method in less than 15 minutes representing a 96% reduction in computational time. A similar pattern could be observed in all other catchments when the computational time was reduced by 97% in Vah@LM and 99% in Hron@Brezno. As expected the highest computational efficiency was observed in the case of the TP method ranging only around 1 minute. Since the accuracy of the TP method is questionable especially in mountainous catchments [9] and the objective of the study was only to accelerate the estimation of areal rainfall using the IDW method the discussion was foccused on the comparison of IDW-t and IDW-a methods

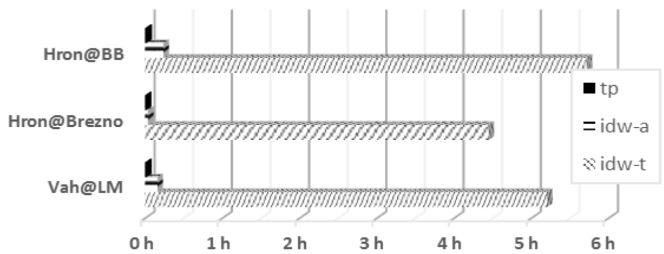

FIG. 6: COMPARISON OF TIME NEEDED FOR ESTIMATION OF AREAL RAINFALL FOR THE WHOLE DATASET WITH MISSING DATA.

TABLE 3
TIME NEEDED FOR CALCULATION OF MEAN DAILY AREAL RAINFALL USING THE THIESSEN POLYGON (TP), TRADITIONAL IDW (IDW-T) AND ACCELERATED IDW (IDW-A) METHOD. CALCULATION FOR THE PERIOD BETWEEN 1961 AND 2010 WITH MISSING DATA.

Catchment	Elapsed time		
	TP	IDW-t	IDW-a
Hron@BB	70s	5h 43m 11s	14m 22s
Hron@Brezno	44s	4h 26m 56s	2m 20s
Vah@LM	59s	5h 12m 17s	9m 43s

One of the task with the highest computational requirements in the IDW-a method is the calculation of weights. This process requires to interpolate synthetic rainfall depths into each cell of the grid into which the catchment is divided. This process is repeated for each rainfall gauge and for each observed combination of missing stations. In the case of a large number of missing data in an analyzed dataset the new IDW-a method could be less efficient and comparable with traditional IDW-t method. In order to test the algorithm in ideal conditions in which the dataset would be complete (no missing data) a new test was conducted. In this test all the three methods were used to estimate areal rainfall over the three catchments for a year with no missing data. The time requirements of individual methods and catchments are displayed in Fig. 7 and listed in Table 4. The results are comparable with those from the previous test when the whole dataset with missing data was used. In all three catchments the time needed for the estimation of areal rainfall using the TP method was around 2 seconds (see Table 4) which is less than 1% of the time needed when using the IDW-t method (reference method). The time requirements for the IDW-a method constituted only for 3 to 7% of the IDW-t method.

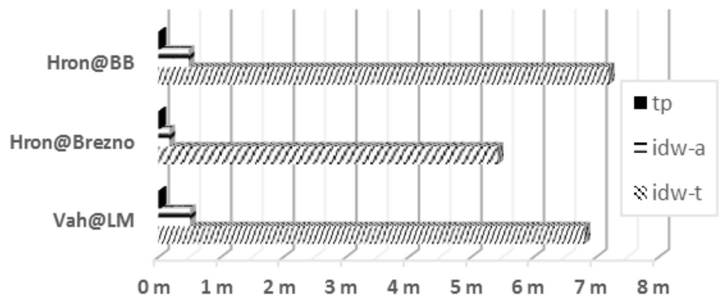

FIG. 7: COMPARISON OF TIME NEEDED FOR ESTIMATION OF AREAL RAINFALL FOR A PERIOD OF 1 YEAR WITHOUT MISSING DATA.

TABLE 4
TIME NEEDED FOR CALCULATION OF MEAN DAILY AREAL RAINFALL USING THE THIESSEN POLYGON (TP), TRADITIONAL IDW (IDW-T) AND ACCELERATED IDW (IDW-A) METHOD. CALCULATION FOR 1 YEAR WITHOUT MISSING DATA.

Catchment	Elapsed time		
	TP	IDW-t	IDW-a
Hron@BB	2s	7min 12s	29s
Hron@Brezno	1.8s	5min 25s	10s
Vah@LM	2s	6min 48s	30s

V. CONCLUSION

The process of estimation areal precipitation is of high importance and is an integral part of solving various tasks in hydrology and other disciplines utilizing mainly conceptual rainfall-runoff models which are not distributed but work as lumped models. The number of methods for estimation of areal rainfall differ mainly in their accuracy but also in computational requirements. One of the methods which proved to be more accurate than the simple method utilizing the Thiessen polygons is an inverse distance interpolation method [8]. However, this method is computationally far less effective than the TP method. The aim of this study was to propose an algorithm which would significantly reduce the time needed for estimation of areal rainfall using the traditional IDW-t method. The new algorithm was tested not only in the terms of computational efficiency but also in the values areal rainfall it estimated. The algorithm was compared with the TP and IDW-

t methods on data from three mountainous catchments situated in the central part of Slovakia. Two tests on two different datasets were performed for each catchment: 1) the whole dataset with missing data and 2) one year with no missing data.

The results showed that when comparing the values of estimated areal rainfall the IDW-a and IDW-t methods give practically identical results with mean absolute error in all three catchments lower than 0.001 (Table 2). As expected the estimates of the TP method when compared to the IDW-t method were slightly different with the maximum absolute error ranging from 3.8 mm (Hron@Brezno) to 7.3 mm (Vah@LM). When comparing the computational efficiency of the methods the new IDW-a method has significantly reduced the time needed for the estimation of areal rainfall using the traditional IDW-t method. In all three catchments the calculation was reduced by more than 96% in the case of analyzing the whole dataset (Table 3) and by more than 92% in the case of analyzing only 1 year (Table 4). In the case of the Hron@Brezno catchment the computational time was reduced by more than 99% when with the traditional IDW-t method it took 4h 26m 56s and with the new IDW-a method only 2m 20s. Similar results were obtained in the remaining catchments.

The significant reduction of computational time in the new IDW-a method could be used in solving many practical tasks. One of them is preparing areal rainfall data for zonal conceptual rainfall-runoff models where for each zone (most often based on altitude) a separate dataset of rainfall depths is prepared. The computational efficiency of the new method enables to divide the catchment into various number of zones and select the best division based on the selected criteria (e.g. comparison of observed and simulated flows, simulation of accumulation and melting of snow cover).

The process of accelerating the IDW method described in this study could be also applied to other linear estimators such as Kriging.

ACKNOWLEDGEMENTS

The authors would like to thank the Slovak science grant agency for funding their research under the contract No. VEGA 0776/13.

REFERENCES

[1] Creutin, J.D., Obled, C. (1982): “Objective Analyses and Mapping Techniques for Rainfall Fields: An Objective Comparison”, Water Resources Research, 18(2), p. 413-431

[2] Hutchinson, M.F. (1998a): “Interpolation of rainfall data with thin plate smoothing splines. Part I: Two dimensional smoothing of data with short range correlation”, J. Geogr. Inform. Decis. Anal., 2, p. 152-167

[3] Hutchinson, M.F. (1998b): “Interpolation of rainfall data with thin plate smoothing splines. Part II: Analysis of topographic dependence”, J. Geogr. Inform. Decis. Anal., 2, p. 168-85

[4] Kemp, K.K. (ed.) (2008): Encyclopedia of Geographic Information Science, Sage Publications, Los Angeles

[5] Lebel, T., Bastin, G., Obled, C., Creutin, J.D. (1987): “On the Accuracy of Areal Rainfall Estimation: A Case Study”, Water resources research, 23(11), p. 2123-2134

[6] Phillips, D.L., Dolph, J., Marks, D. (1992): “A comparison of geostatistical procedures for spatial analysis of precipitations in mountainous terrain”, Agr. Forest. Meteorol., 58, p. 119-41

[7] Sevruk, K. (2005): “Rainfall Measurement: Gauges”, In: Anderson, M.G. (ed), Encyclopedia of Hydrological Sciences, Wiley, Chichester, p. 529-535

[8] Tabios, G.Q., Salas, J.D. (1985): “A comparative analysis of techniques for spatial interpolation of precipitation”, Water Resources Bulletin, 23(3), p. 365-380

[9] Taesombat, W., Sriwongsitanon, N. (2009): “Areal rainfall estimation using spatial interpolation techniques”, ScienceAsia, 35, p. 268-275

[10] Thiessen, A. H. (1911): “Precipitation averages for large areas”, Monthly Weather Rev., 39, p. 1082-1084